Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE @ DIRECT•

An International Journal
computers & mathematics
with applications

ELSEVIER Computers and Mathematics with Applications 46 (2003) 1103–1110
www.elsevier.com/locate/camwa

A Structural Approximation Method to Generate the Optimal Auto-Sleep Schedule for Computer Systems

H. OKAMURA AND T. DOHI
Department of Industrial and Systems Engineering
Hiroshima University
4-1 Kagamiyama 1 Chome, Higashi-Hiroshima 739-8527, Japan
<okamu><dohi>@rel.hiroshima-u.ac.jp

S. OSAKI
Faculty of Mathematical Sciences and Information Engineering
Nanzan University
27 Seirei-cho, Seto-shi, Aichi, 489-0863, Japan
shunji@it.nanzan-u.ac.jp

Abstract—This paper addresses a problem of energy reduction on computer systems, namely the problem of determining the optimal auto-sleep schedule on which the system is changed into a low-power state. For this problem, we provide a stochastic model for a computer system with auto-sleep function. In our modeling framework, user requests arrive at the system in accordance with a general renewal process and are served by generally distributed service times. To obtain the optimal auto-sleep schedule minimizing the expected power consumption per unit time in the steady state, the phase-type approximation is applied to the renewal arrival process. In numerical examples, we investigate accuracy of the phase-type approximation through a simulation study. © 2003 Elsevier Ltd. All rights reserved.

Keywords—Auto-sleep schedule, Power saving, Renewal process, Phase-type distribution, EM algorithm.

1. INTRODUCTION

Recently, the auto-sleep function of a hard disk system or a display in computer systems was recognized to be important in terms of power management. The auto-sleep function is installed in most computer systems as a standard function. The optimal design for the auto-sleep schedule is an important problem, because, after waking up from sleep mode with low-power, the system undergoes not only the performance degradation, but also wastage for power consumption. In other words, the excessive shutdown for the system may cause more waste power consumption. In particular, the problem is critical for laptop computers with limited capacity of battery.

Sandoh, Hirakoshi and Kawai [1] consider the optimal design problem for the auto-sleep function in a hard disk system. Dohi, Kaio and Osaki [2] propose a statistical nonparametric method

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B), 13480109 (2001–2003), and Nanzan University Pache Research Subsidy I-A.

 Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
doi: 10.1016/S0898-1221(03)00316-X

to estimate the optimal sleep schedule. However, these seminal works simplify the underlying problem extremely, and therefore, their models do not adequately represent the stochastic behavior of computer systems with auto-sleep function. More valid formulation is made by Okamura, Dohi and Osaki [3,4]. They consider two kinds of models (Type I and Type II models) with and without cancellation of user requests, respectively. More precisely, Type I model assumes that the arrival of user requests while the system is busy are refused. On the other hand, in Type II model, the system has an infinite buffer and receives all the user requests. If the system is busy, the user requests are stored in the buffer and are served exhaustively. Type I and II models correspond to single-task and multitask systems, respectively. Okamura, Dohi and Osaki [3,4] show that the optimal strategies on the auto-sleep schedule for both models are the *switching strategies* when the user requests follow the homogeneous Poisson process. In the switching strategy, the system, when it is idle, always either sleeps or operates. Also, Okamura, Dohi and Osaki [3,4] apply the simple parametric approximation methods by Miyazawa [5] and the diffusion approximation to computing the optimal auto-sleep strategies approximately in the renewal arrival case. It should be noted, however, that the approximate auto-sleep schedule cannot guarantee sufficient accuracy in practical application.

In this paper, we propose the phase-type approximation method to generate the approximate auto-sleep schedule in the general renewal arrival case. In particular, this paper considers Type I model, namely, the single-task system. The phase-type approximation is based on the phase-type renewal process whose interarrival time distribution is the phase-type distribution, and is effective to approximate general renewal processes. In fact, Altiok [6] and Heijden [7] show that the phase-type distribution can approximate arbitrary probability distributions. Asmussen and Koole [8] also prove that the phase-type renewal process is weakly dense in the class of stationary simple point processes. Thus, by applying the phase-type approximation to the design of the auto-sleep function, the approximate auto-sleep schedule is expected to be more accurate and more practical than that in [3,4].

The paper is organized as follows. Section 2 describes the stochastic model for a computer system with auto-sleep function, and gives an implicit formula of the expected power consumption per unit time in the steady state. In Section 3, the phase-type approximation is introduced to approximate the request arrival process, where two statistical estimation methods for the phase-type distribution are developed. Section 4 is devoted to investigate the accuracy of approximation methods proposed in Section 3. Finally, we conclude this paper with some remarks in Section 5.

2. MODEL DESCRIPTION

2.1. Notation and Assumptions

Suppose that user requests follow a renewal process. The interval times between the $(k-1)^{\text{th}}$ and the k^{th} arrivals are denoted by X_k, which are nonnegative i.i.d. random variables having the distribution function $F(t)$ with mean $1/\lambda$ (> 0) and variance σ_a (> 0). The k^{th} user request needs the processing time S_k, which is a nonnegative random variable having the distribution function $H(t)$ with mean $1/\mu$ (> 0) and variance σ_s (> 0). Define the following four system states.

Busy: The system is processing user requests, where set-up time τ (> 0) is needed before starting to process each request. After completion of processing the request, the system becomes idle. In the busy state, the power consumption per unit time is P_1 (> 0).

Idle: The system is waiting for an additional user request. If the additional request arrives before the total time spent in the idle state becomes t_0, the system begins processing the user request with the set-up time τ. Otherwise, when the total time spent in idle state is t_0, the system goes to sleep state. In this paper, the time t_0 is called the *auto-sleep schedule*. The power consumption per unit time in the idle state is also P_1 (> 0) to simplify the analysis.

Sleep: The system is sleeping to reduce the energy consumption. For convenience, we assume that the power consumption in sleep state is zero. After a user request arrives, the system wakes up and goes to warm-up state immediately.

Warm-up: The system wakes up from sleep state. Since the warm-up state is transient, the system goes to busy state after elapsing the time s. In the warm-up state, the power consumption per unit time is P_2 ($> P_1$).

In this paper, the user requests arriving while the system is busy are refused. Figure 1 depicts the possible behavior of the computer system with auto-sleep function.

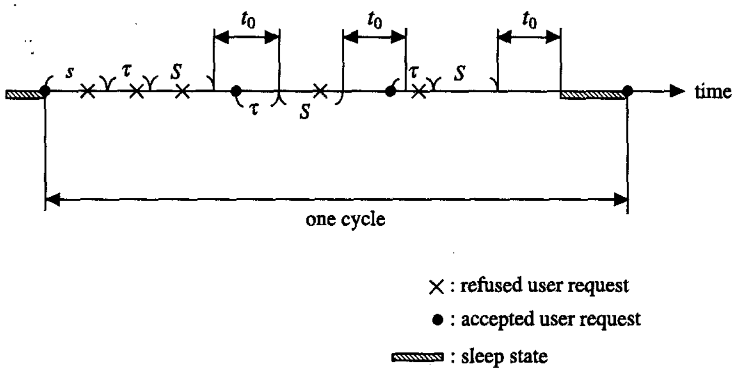

Figure 1. Possible behavior of a computer system with auto-sleep function.

2.2. Formulation

Let us consider the expected power consumption per unit time in the steady state as a criterion to evaluate energy efficiency on the auto-sleep function.

Define the residual life by γ_t having the distribution function $I(x \mid t)$, where the subscript t denotes the elapsed time. In our modeling framework, since the user requests occur during the busy state, the residual life of the arrival process represents the time length of idle state. Let $M(t)$ be the renewal function of the arrival process. The residual life distribution is then given by

$$I(x \mid t) = F(t+x) - \int_0^t \bar{F}(t+x-y)\, dM(y), \tag{1}$$

where, in general, $\bar{\psi}(\cdot) = 1 - \psi(\cdot)$ is a survivor function.

Define the time period from the beginning of warm-up state to the next as one cycle. Then, we have the mean time length of one cycle

$$T(t_0) = s + \tau + \frac{1}{\mu} + \int_0^\infty \mathrm{E}[\gamma_{s+\tau+x}]\, dH(x) + \mathrm{E}[N]\left(\tau + \frac{1}{\mu} + \int_0^\infty \mathrm{E}[\gamma_{\tau+x}]\, dH(x)\right), \tag{2}$$

where $\mathrm{E}[N]$ is the expected number of transitions from idle to busy during one cycle. The probability mass function of N is given by

$$\Pr\{N=n\} = \int_0^\infty I(t_0 \mid s+\tau+x)\bar{I}(t_0 \mid \tau+x) I(t_0 \mid \tau+x)^{n-1}\, dH(x), \qquad \text{for } n = 1, 2, \ldots. \tag{3}$$

Hence, the expected number of transitions from idle to busy during one cycle becomes

$$\mathrm{E}[N] = \frac{\Pr\left\{\int_0^\infty \gamma_{s+\tau+x} \le t_0\right\} dH(x)}{\Pr\left\{\int_0^\infty \gamma_{\tau+x} > t_0\right\} dH(x)} = \frac{\int_0^\infty I(t_0 \mid s+\tau+x)\, dH(x)}{\int_0^\infty \bar{I}(t_0 \mid \tau+x)\, dH(x)}. \tag{4}$$

In a fashion similar to the mean time length of one cycle, the expected power consumption during one cycle is expressed in the form

$$\begin{aligned} C(t_0) = \left\{\frac{\rho}{1-\rho}P_1 + P_2\right\}s + \frac{P_1\tau}{1-\rho} + P_1\left\{\frac{\rho}{1-\rho}\mathrm{E}[\eta_{s+\tau}] + \mathrm{E}[\eta_{s+\tau}\wedge t_0]\right\} \\ + \mathrm{E}[N]\left\{\frac{P_1\tau}{1-\rho} + P_1\left(\frac{\rho}{1-\rho}\mathrm{E}[\eta_\tau] + \mathrm{E}[\eta_\tau \wedge t_0]\right)\right\}, \end{aligned} \tag{5}$$

where

$$\mathrm{E}[\eta_t \wedge t_0] = \mathrm{E}[\min(\eta_t, t_0)] = \int_0^{t_0} u\, dI(u \mid t) + t_0 \bar{I}(t_0 \mid t). \tag{6}$$

From the renewal reward theorem, we obtain the expected power consumption per unit time in the steady state,

$$V(t_0) = \frac{C(t_0)}{T(t_0)}. \tag{7}$$

Then the problem is to find the optimal auto-sleep schedule t_0^* which minimizes the expected power consumption per unit time in the steady state $V(t_0)$.

3. THE PHASE-TYPE APPROXIMATION

3.1. Approximate Formula

In general, the expected power consumption is difficult to formulate explicitly in the general renewal arrival case. This is due to an analytical difficulty to express the renewal function. In spite of getting the explicit form, we provide a structural approximation scheme to generate the optimal auto-sleep schedule. This scheme is based on the approximation of the general renewal arrival process by the phase-type renewal process, so that we get an approximate form of the residual distribution $I(t \mid x)$ by the phase-type distribution.

Before developing the phase-type approximation, we describe the phase-type renewal process. Consider a Markov process on state space $\{1, 2, \ldots, m+1\}$, where $\{1, 2, \ldots, m\}$ denote transient states called *phases*, and $\{m+1\}$ means the absorbing one. The initial probability vector for the Markov process is given by $(\boldsymbol{\alpha}, 0)$, where $\boldsymbol{\alpha}$ is $1 \times m$ probability vector. The behavior of the stochastic process is similar to the Markov process with an infinitesimal generator T until absorption in state $m+1$, where T is a matrix with components λ_{ij} (> 0), $1 \le i, j \le m$, $j \ne i$, and $-\lambda_{ii}$ (< 0). When we treat the phase-type renewal process, an absorption event can be regarded as an arrival of user requests. The process is restarted with the phase following the initial probability vector after the absorption event. The time interval between successive arrivals can be represented by the phase-type distribution with parameter $(\boldsymbol{\alpha}, T)$, where the interarrival time distribution is given by $F_{PH}(t) = 1 - \boldsymbol{\alpha}\exp(Tt)\mathbf{e}$ with a column vector of 1s, $\mathbf{e}$.

Let us now return our argument to the phase-type approximation. The stochastic processes, N_t and J_t, denote the number of arrivals in $(0, t]$ and the phase at time t, respectively. We also define the matrix $P(n, t)$ with components $P_{ij}(n, t)$, which is the transition probability given by

$$P_{ij}(n, t) = \Pr\{N_t = n,\ J_t = j \mid N_0 = 0,\ J_0 = i\}. \tag{8}$$

Then, Kolmogorov's forward equation is given by

$$\begin{aligned} \frac{d}{dt}P(0, t) &= P(0, t)T, \\ \frac{d}{dt}P(n+1, t) &= P(n+1, t)T + P(n, t)\mathbf{T^0}\boldsymbol{\alpha}, \qquad \text{for } n = 1, 2, \ldots, \\ P(0, 0) &= I, \quad P(n, 0) = O, \qquad \text{for } n = 1, 2, \ldots, \end{aligned} \tag{9}$$

where $\mathbf{T^0} = -T\mathbf{e}$ is the column vector, I is an identity matrix, and O is a zero matrix. Let $P^*(z,t) = \sum_{n=0}^{\infty} P(n,t)z^n$ be the matrix generating function. From equation (9), we obtain

$$P^*(z,t) = \sum_{n=0}^{\infty} P(n,t)z^n = \exp\left\{\left(T + z\mathbf{T^0}\boldsymbol{\alpha}\right)t\right\}. \tag{10}$$

Hence, we can derive a probability vector $\mathbf{g}(t)$ with component $g_j(t)$ which means the probability that the phase at time t is j, that is,

$$\mathbf{g}(t) = \boldsymbol{\alpha}\exp\left\{\left(T + \mathbf{T^0}\boldsymbol{\alpha}\right)t\right\}. \tag{11}$$

From the Markov property of the phase-type renewal process, the residual life distribution can be written as

$$I_{PH}(x \mid t) = 1 - \mathbf{g}(t)\exp(Tx)\mathbf{e}. \tag{12}$$

Therefore, the residual life distributions in equations (2) and (5) are replaced by the above phase-type distribution, namely $I(x \mid t) \approx I_{PH}(x \mid t)$. To this end, we derive an approximate formula of the expected power consumption per unit time in the steady state.

3.2. Statistical Estimation Procedures

Since the phase-type renewal process is usually composed of two kinds of stochastic processes, observable and unobservable processes, we cannot apply the usual statistical parameter estimation methods, such as the maximum likelihood method. Thus, we introduce two statistical methods for estimating the parameter of the phase-type distribution.

(i) *The moment matching.*

Heijden [7] proposes the moment matching conditions. If there are n unknown parameters, they are determined by fitting the first n moments to the sample moments estimated from real data. If the interarrival time distribution of the phase-type renewal process obeys the following Coxian-2 distribution:

$$T = \begin{bmatrix} -\lambda_1 & 0 \\ \lambda_2 & -\lambda_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha} = (1-a, a), \tag{13}$$

the estimates are given by

$$a = \frac{\lambda_2}{\lambda_1}(m_1\lambda - 1), \tag{14}$$

$$\lambda_1 = \frac{3m_1m_2 - m_3 - \sqrt{m_3^2 + 18m_2^3 + 24m_1^3m_3 - 9m_1m_2(3m_1m_2 + 2m_3)}}{3m_2^2 - 2m_1m_3}, \tag{15}$$

and

$$\lambda_2 = \frac{2(m_1\lambda_1 - 1)}{m_2\lambda_1 - 2m_1}, \tag{16}$$

where m_1, m_2, and m_3 are the first three moments of the interarrival time.

(ii) *The EM algorithm for phase-type distribution.*

The EM (expectation-maximization) algorithm is an iterative method for the maximum likelihood estimation [9,10]. It is useful to parameterize statistical models including the incomplete data. Suppose that $Y = u(X)$ is observed and that X is unobserved, where Y and X have the probability density functions g_γ and f_γ, respectively. Then, the $(n+1)^{\text{th}}$ step in the EM algorithm is to find the value γ_{n+1} which maximizes

$$\gamma_{n+1} = \underset{\gamma}{\operatorname{argmax}}\, \mathrm{E}[\log f_\gamma(X) \mid u(X) = y;\ \gamma_n], \tag{17}$$

where y is the observed data and γ_n is the current estimated parameter set after the n^{th} step in the EM algorithm (see, e.g., [11]). In particular, when the interarrival time distribution has the phase-type distribution, the EM algorithm is described as follows.

Let $(y_1, y_2, \ldots, y_n)$ be the observed sample data. Then the $(k+1)^{\text{th}}$ iteration of the EM algorithm is given by the following.

E-STEP. Calculate

$$\pi_i^{(k+1)} = \sum_{l=1}^{n} \mathrm{E}\left[\pi_i^{(k)} \mid y_l;\ \hat{\alpha}^{(k)}, \hat{T}^{(k)}\right], \qquad \text{for } i = 1, \ldots, m, \tag{18}$$

$$\xi_i^{(k+1)} = \sum_{l=1}^{n} \mathrm{E}\left[\xi_i^{(k)} \mid y_l;\ \hat{\alpha}^{(k)}, \hat{T}^{(k)}\right], \qquad \text{for } i = 1, \ldots, m, \tag{19}$$

$$\Lambda_{ij}^{(k+1)} = \sum_{l=1}^{n} \mathrm{E}\left[\Lambda_{ij}^{(k)} \mid y_l;\ \hat{\alpha}^{(k)}, \hat{T}^{(k)}\right], \qquad \text{for } i \neq j,\ i = 1, \ldots, m \text{ and } j = 1, \ldots, m. \tag{20}$$

M-STEP. Generate the new estimates,

$$\hat{\alpha}_i^{(k+1)} = \frac{\pi_i^{(k+1)}}{n}, \qquad \hat{t}_{ij}^{(k+1)} = \frac{\Lambda_{ij}^{(k+1)}}{\xi_i^{(k+1)}}, \qquad \hat{t}_{ii}^{(k+1)} = -\left(\frac{\Lambda_{i0}^{(k+1)}}{\xi_i^{(k+1)}} + \sum_{j=1, j\neq 1}^{m} \hat{t}_{ij}^{(k+1)}\right), \tag{21}$$

where $\hat{\alpha}_i$ and $\hat{t}_{ij}$ are the elements of $\hat{\alpha}$ and $\hat{T}$, respectively. In the above formulas, π_i is the number of Markov processes starting at the state i, ξ_i is the total time spent in the state i, and Λ_{ij} is the total number of jumps from the state i to j.

4. NUMERICAL EXAMPLES

In this section, we investigate the accuracy of the phase-type approximation proposed in Section 3. Suppose that user requests follow the renewal process with interarrival time distribution $F(t) = 1 - \exp\{-(t/\beta_a)^{m_a}\}$. The shape and scale parameters of the above distribution are set as $m_a = 0.5$ and $\beta_a = \rho/\Gamma(1+1/m_a)$, respectively, where $\Gamma(\cdot)$ is the standard gamma function. We assume that the processing time distribution is the exponential distribution, $H(t) = 1 - \exp(-t)$, and the interarrival time distribution is the Coxian-2 distribution. To perform the sensitivity analysis, the varying traffic intensity ρ from 0.1 to 0.9 as well as $P_1 = 1.0$, $P_2 = 4.0$, $\tau = 0.1$, and $s = 1.0$ are assumed.

In particular, to compare the phase-type approximation with other approximation, we calculate the optimal auto-sleep schedule based on the equilibrium approximation [4]. In this approximation scheme, the residual life distribution can be approximated by the equilibrium distribution,

$$I(t \mid x) \approx F_e(t) = \lambda \int_0^t \bar{F}(u)\, du. \tag{22}$$

Tables 1 and 2 present the optimal auto-sleep schedule and their associated minimum expected power consumptions per unit time in the steady state for respective approximation schemes. In the phase-type approximation, we apply both the moment matching and the EM algorithm to estimating the model parameters (see Table 2).

To evaluate the performance on approximation, the Monte Carlo simulation is performed and we calculate the simulated values on the power consumption under the approximate optimal auto-sleep schedule. The resulting values in brackets on each table indicate the lower and upper bounds on the confidence interval with significant level 95% for the simulated power consumption.

From these results, it can be observed that the expected power consumptions based on the equilibrium approximation and the moment matching are quite different from the simulated values. On the other hand, the estimated power consumptions by the EM algorithm are within the confidence intervals with significant level 95%, and thus, the phase-type approximation with EM algorithm estimates the expected power consumption accurately. However, in terms of determination of the optimal sleep schedule, it is noted that the EM algorithm does not always give

Table 1. Optimal auto-sleep schedules based on the equilibrium approximation.

ρ	$\hat{t}_0^*$	$\hat{V}(t_0^*)$	Simulation
0.1	0.000	0.159	0.299 (0.278, 0.320)
0.2	0.000	0.298	0.484 (0.454, 0.515)
0.3	0.006	0.421	0.617 (0.579, 0.655)
0.4	0.189	0.529	0.762 (0.728, 0.797)
0.5	0.476	0.620	0.804 (0.761, 0.846)
0.6	0.806	0.695	0.871 (0.825, 0.918)
0.7	1.154	0.756	0.896 (0.855, 0.937)
0.8	1.514	0.806	0.952 (0.911, 0.994)
0.9	1.874	0.847	0.918 (0.873, 0.962)

Table 2. Optimal auto-sleep schedules based on the phase-type approximations.

	Moment Matching			EM Algorithm		
ρ	$\hat{t}_0^*$	$\hat{V}(t_0^*)$	Simulation	$\hat{t}_0^*$	$\hat{V}(t_0^*)$	Simulation
0.1	0.000	0.379	0.299 (0.278, 0.320)	0.000	0.334	0.299 (0.278, 0.320)
0.2	0.000	0.608	0.484 (0.454, 0.515)	0.000	0.622	0.484 (0.454, 0.515)
0.3	3.128	0.705	0.665 (0.626, 0.704)	0.000	0.644	0.617 (0.579, 0.655)
0.4	3.265	0.747	0.790 (0.762, 0.819)	0.037	0.852	0.773 (0.736, 0.809)
0.5	3.058	0.778	0.815 (0.780, 0.851)	0.421	0.922	0.808 (0.764, 0.851)
0.6	2.826	0.805	0.861 (0.824, 0.899)	∞	1.000	0.997 (0.991, 1.003)
0.7	2.625	0.828	0.881 (0.848, 0.914)	∞	1.000	1.000 (1.000, 1.000)
0.8	2.454	0.850	0.942 (0.906, 0.977)	∞	1.000	1.000 (1.000, 1.001)
0.9	2.310	0.870	0.918 (0.880, 0.956)	0.797	0.710	0.964 (0.908, 1.019)

the accurate estimates. By contrast, the equilibrium approximation and the moment matching are more useful for estimating the optimal auto-sleep schedule.

5. CONCLUDING REMARKS

In this paper, we have considered the stochastic auto-sleep model under renewal arrival processes, and have proposed the phase-type approximation to generate approximately the optimal auto-sleep schedule minimizing the expected power consumption per unit time in the steady state. In numerical examples, we have compared the proposed method with the existing one quantitatively, and investigated usefulness of the proposed methods. As a result, we have shown that the phase-type approximation could be effective for finding the optimal auto-sleep schedule and estimating the power consumption.

REFERENCES

1. H. Sandoh, H. Hirakoshi and H. Kawai, An optimal time to sleep for an auto-sleep system, *Computers & Operations Research* **23**, 221–227, (1996).
2. T. Dohi, N. Kaio and S. Osaki, Nonparametric approach to power-saving strategies for a portable personal computer, *Electronics and Communications in Japan, Part 3* **80**, 80–90, (1997).
3. H. Okamura, T. Dohi and S. Osaki, On the effects of power saving by auto-sleep functions on a computer system I—Modeling by a renewal process, (in Japanese), *Transactions on IPSJ* **39**, 1858–1869, (1999).
4. H. Okamura, T. Dohi and S. Osaki, On the effects of power saving by auto-sleep functions on a computer system II—Queueing models, (in Japanese), *Transactions on IPSJ* **40**, 1027–1040, (2000).
5. M. Miyazawa, A generalized Pollaczek-Khintchine formula for the $GI/GI/1/K$ queue and its application to approximation, *Stochastic Models* **3**, 53–65, (1989).
6. T. Altiok, On the phase-type approximations of general distributions, *IIE Transactions* **17**, 110–116, (1985).
7. M.C. Heijden, On the three-moment approximation of a general distribution by a Coxian distribution, *Probability in the Engineering and Informational Sciences* **2**, 257–261, (1988).

8. S. Asmussen and G. Koole, Marked point processes as limits of Markovian arrival streams, *Journal of Applied Probability* **30**, 356–372, (1993).
9. A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* **39**, 1–38, (1977).
10. C.F.J. Wu, On the convergence properties of the EM algorithm, *Annals of Statistics* **11**, 95–103, (1983).
11. S. Asmussen, O. Nerman and M. Olsson, Fitting phase-type distributions via the EM algorithm, *Scandinavian Journal of Statistics* **23**, 419–441, (1996).